Some Monotonicity Properties of Schur Powers of Matrices and Related Inequalities*

Samuel Karlin
Department of Mathematics
Stanford University
Stanford, California 94305

and

Friedemann Ost
Department of Mathematics
Stanford University
Stanford, California 94305
and
Department of Applied Mathematics and Statistics
Technical University of Munich
Arcisstr aβe 21
D-8000 Munich 2, Germany

Submitted by Hans Schneider

ABSTRACT

A series of inequalities are developed relating the spectral radius $\rho(A \circ B)$ of the Schur product $A \circ B$ of two nonnegative matrices A and B with those of $\rho(A \circ A)$ and $\rho(B \circ B)$ yielding $\rho(A \circ B) \leqslant [\rho(A \circ A)\rho(B \circ B)]^{1/2}$. As a corollary it is proved that the spectral radius of the Schur powers $\rho_r = \rho(A^{[r]})$, $A^{[r]} = A \circ A \circ \cdots \circ A$ (r factors) satisfies $(1/r)\log \rho_r$ is decreasing while $(1/r - 1)\log \rho_r$ is increasing, the latter provided A is a stochastic matrix. The entropy of a finite stationary Markov chain is identified with $d\rho_r/dr|_{r=1}$. A number of majorization comparisons for the spectral radius of Schur powers is given.

*Supported in part by NIH Grant 5R01 GM10452-20, NSF Grant MSC79-24310, and a NATO Fellowship by the DAAD.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/85/$3.30

1. INTRODUCTION

The Schur product of the matrices $A = \|a_{ij}\|$ and $B = \|b_{ij}\|$, denoted by $A \circ B = \|a_{ij}b_{ij}\|$ (also called its Hadamard product), is playing an increasing role in many areas of linear algebra, analysis, and multivariate probabalistic and statistical analysis (e.g., see [9, 2, 10]).

Let $P = \|p_{ij}\|$ be an $m \times m$ nonnegative irreducible matrix, and let

$$P^{[r]} = P \circ P \circ \cdots \circ P$$

denote the rth Schur power of P with itself. It is convenient to consider $P^{[r]} = \|(p_{ij})^r\|$ for all positive r, and if p_{ij} are all positive, then $P^{[r]}$ is defined for all real r. The Frobenius theory of positive matrices tells us that the spectral radius ρ_r of $P^{[r]}$ is the eigenvalue of maximal magnitude having unique (modulo a scalar) strictly positive right and left associated eigenvectors.

The results described in this paper were motivated by studies of comparing homologies and word relationships between random sequences generated from an m-letter alphabet $\mathscr{A}$ where the successive letters in each sequence $\mathscr{S}$ occur as independent realizations of an m-state Markov chain governed by the transition matrix P.

With such a randomly generated sequence $\mathscr{S}$ of total length N, consider $L_r^{(N)}$ determined as the length of the longest word (a word in $\mathscr{S}$ of length k is a contiguous set of k letters from $\mathscr{A}$) occurring at least r times in $\mathscr{S}$. We have established [4] that the expected length of $L_r^{(N)}$ is of asymptotic order

$$\frac{\log\binom{N}{r}}{(-\log \rho_r)}, \qquad \text{where} \quad \rho_r = \rho(P^{[r]}).$$

Another natural example leading to Schur products involves comparing between several sequences. Consider two random strings of N letters, $\mathscr{S}_1$ and $\mathscr{S}_2$ (from the alphabet $\mathscr{A}$), the realization of $\mathscr{S}_1$ governed by the Markov transition matrix P and the realization of $\mathscr{S}_2$ governed by the Markov transition matrix Q. We assume $\mathscr{S}_1$ and $\mathscr{S}_2$ are generated independently. Let

$$W_{P,Q}^{N}$$

be the length of the longest word common to $\mathscr{S}_1$ and $\mathscr{S}_2$. Of particular interest (for biological reasons) is the case having $Q = P^*$ such that the transition matrix Q is that of the time reversed Markov chain to P. Another important class of examples pertains to the stipulation of $Q = \Pi^{-1} P \Pi$ where Π is a fixed permutation matrix.

We have proved that the random variable $W_{P,Q}^N$ grows on the average as $(\log N^2)/(-\log \rho_{P,Q})$, where $\rho_{P,Q} = \rho(P \circ Q)$.

The foregoing probabalistic models lead to natural comparison inequalities on Schur products, set forth next in Section 2.

2. COMPARISONS ON THE SPECTRAL RADIUS OF SCHUR PRODUCTS

It is convenient to recall several standard facts pertaining to the spectral radius of positive matrices.

The following characterizations of the spectral radius for nonnegative matrices will serve. Let C be a nonnegative matrix. Then

$$\rho(C) = \sup_{\mu \in \Delta} \mu, \tag{2.1}$$

where $\Delta = \{\mu | \mu \text{ real and } C\mathbf{x} \geqslant \mu \mathbf{x} \text{ for some } \mathbf{x} \geqslant \mathbf{0},\ \mathbf{x} \neq \mathbf{0}\}$. (The notation $\mathbf{z} \geqslant \mathbf{w}$ signifies that $\mathbf{z} - \mathbf{w}$ is a nonnegative vector.)

If there exists $\mathbf{y} \gg \mathbf{0}$ (strictly positive) satisfying $C\mathbf{y} \leqslant \gamma \mathbf{y}$, then

$$\rho(C) \leqslant \gamma. \tag{2.2}$$

An elementary deduction from (2.1) entails for two nonnegative matrices, B dominating componentwise A, $A \leqslant B$, that $\rho(A) \leqslant \rho(B)$. This confirms the fact

$$\rho(C) \geqslant \rho(\tilde{C}), \tag{2.3}$$

where $\tilde{C}$ is a principal submatrix of C; and if C is irreducible nonnegative, then strict inequality obtains in (2.3).

It is familiar (e.g., see [3]) that $A \circ B$ is a principal submatrix for the Kronecker product matrix $A \otimes B$. This property implies that for A and B nonnegative matrices,

$$\rho(A \circ B) \leqslant \rho(A)\rho(B). \tag{2.4}$$

Indeed, by (2.3) we have the first inequality of

$$\rho(A \circ B) \leqslant \rho(A \otimes B) = \rho(A)\rho(B),$$

the last equation resulting because the set of eigenvalues of $A \otimes B$ comprise the products of all eigenvalues of A with those of B.

We sharpen the inequality (2.4) to

THEOREM 2.1. *Let A and B be nonnegative matrices. Then*

$$\rho(A \circ B) \leqslant \sqrt{\rho(A \circ A)}\,\sqrt{\rho(B \circ B)}\,. \tag{2.5}$$

When A and B are also irreducible, then equality occurs if and only if $A = DBD^{-1}$ where D is a diagonal positive matrix. If A and B are stochastic, that is, $A\mathbf{u} = \mathbf{u}$ and $B\mathbf{u} = \mathbf{u}$ ($\mathbf{u} = (1, \dots, 1)$ exhibiting only unit coordinates) and equality holds in (2.5), *then $D = I$ = identity matrix.*

Proof. Assume that A and B are each an irreducible nonnegative matrix. Let $\mathbf{x} = (x_1, \dots, x_m)$ be the unique (apart from a scalar constant) positive eigenvector for $A^{[2]}$, so that

$$(A \circ A)\mathbf{x} = \rho(A^{[2]})\mathbf{x}, \tag{2.6}$$

and $\mathbf{y} = (y_1, y_2, \dots, y_m)$ similarly corresponds to $B^{[2]}$ satisfying

$$(B \circ B)\mathbf{y} = \rho(B^{[2]})\mathbf{y}. \tag{2.7}$$

Let $\mathbf{z} = (z_1, \dots, z_m)$ be the positive vector with components $z_i = \sqrt{x_i y_i}$. Applying Schwartz's inequality, we have

$$\sum_{j=1}^{m} a_{ij} b_{ij} \sqrt{x_j y_j} \leqslant \sqrt{\sum_{j=1}^{m} a_{ij}^2 x_j}\,\sqrt{\sum_{j=1}^{m} b_{ij}^2 y_j} \tag{2.8}$$

with equality if and only if

$$a_{ij}\sqrt{x_j} = \gamma_i b_{ij}\sqrt{y_j} \qquad \text{for all } i, j. \tag{2.9}$$

In the case of (2.9), squaring both sides and adding over j gives

$$\rho(A^{[2]})x_i = \sum_j a_{ij}^2 x_j = \gamma_i^2 \sum_j b_{ij}^2 y_j = \gamma_i^2 \rho(B^{[2]}) y_i. \tag{2.10}$$

Then $\sqrt{x_i} = c\gamma_i\sqrt{y_i}$ for all i (c is a positive constant) and (2.9) reduces to

$$cA = \Gamma B \Gamma^{-1}, \tag{2.11}$$

where Γ is a positive diagonal matrix having the components of $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ down the diagonal. When A and B are further stochastic matrices (i.e., $A\mathbf{u} = B\mathbf{u} = \mathbf{u}$) and (2.11) holds, we get (with $1/\boldsymbol{\gamma}$ the vector components $1/\gamma_i$)

$$B\frac{1}{\boldsymbol{\gamma}} = c\frac{1}{\boldsymbol{\gamma}},$$

which compels c to be 1 and $\boldsymbol{\gamma} = \mathbf{u} = (1, \ldots, 1)$, so that in this case, with A and B both irreducible stochastic matrices, equality in (2.9) is possible only when $A = B$.

Returning to (2.8) without (2.9), inserting (2.7) and (2.6), we obtain

$$\sum_{j=1}^{m} a_{ij} b_{ij} z_j \leqslant \left(\sqrt{\rho(A^{[2]})\rho(B^{[2]})}\right) z_i$$

with strict inequality applying for at least one component of (2.9). When $C = A \circ B$ is irreducible, since $\mathbf{z}$ is a strictly positive vector, the criterion of (2.2) for the spectral radius assures the conclusion

$$\rho(A \circ B) \leqslant \sqrt{\rho(A^{[2]})\rho(B^{[2]})}$$

with *strict inequality unless*

$$A = DBD^{-1} \tag{2.12}$$

where D is a positive diagonal matrix. By standard continuity arguments the inequality (2.5) extends to any two nonnegative matrices A and B.

If $C = A \circ B$ is reducible, then $\rho(C) = \rho(C_1) = \rho(A_1 \circ B_1) \leqslant \sqrt{\rho(A_1 \circ A_1)\rho(B_1 \circ B_1)}$ for some appropriate submatrices A_1, B_1, and C_1. As $A \circ A$ is irreducible, $\rho(A_1 \circ A_1) < \rho(A \circ A)$, and similarly $\rho(B_1 \circ B_1) <$

$\rho(B \circ B)$ and the inequality sign holds in (2.5). Note if B is reducible, then $A \circ B$ is also reducible. Thus, if $\rho(A), \rho(B) > 0$ and either A or B is irreducible and C reducible, then the inequality sign holds in (2.5). ■

REMARK 1. We can extend the result of Theorem 2.1 as follows. Let A_i be nonnegative matrices and $\alpha_i > 0$ with $\sum_{i=1}^{t} 1/\alpha_i = 1$. Then

$$\rho\left(A_1^{[1/\alpha_1]} \circ A_2^{[1/\alpha_2]} \circ \cdots \circ A_t^{[1/\alpha_t]}\right) \leqslant \prod_{i=1}^{t} \left[\rho(A_i)\right]^{1/\alpha_i},$$

and replacing A_i by $A_i^{[\alpha_i]}$, this reads

$$\rho(A_1 \circ A_2 \circ \cdots \circ A_t) \leqslant \prod_{i=1}^{t} \left[\rho\left(A_i^{[\alpha_i]}\right)\right]^{1/\alpha_i}. \tag{2.13}$$

REMARK 2. For A a stochastic matrix whose elements provide the transition probabilities of a Markov chain, the reversed stochastic process observing the changes of state backwards in time is governed by the stochastic matrix $D_{\boldsymbol{\pi}}^{-1} A' D_{\boldsymbol{\pi}}$, where $D_{\boldsymbol{\pi}}$ is the diagonal matrix with components of $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ down the diagonal, and $\boldsymbol{\pi} A = \boldsymbol{\pi}$, so that $\boldsymbol{\pi}$ is the stationary frequency vector of A. Here A' is the transpose of A.

The following corollary of Theorem 2.1 is motivated in part by the probabilistic models described in Section 1.

COROLLARY 2.1. *Let A be a nonnegative matrix and S any permutation matrix. Then*

$$\rho(A \circ A') \leqslant \rho(A \circ A) \tag{2.14a}$$

and

$$\rho(A \circ S'AS) \leqslant \rho(A \circ A), \tag{2.14b}$$

where A' is the transpose matrix to A.

When A is an irreducible Markov matrix, equality occurs in (2.14a) *if and only if $A' = DAD^{-1}$, where the vector $\mathbf{d}$ of diagonal elements in D satisfies $\mathbf{d}A = \mathbf{d}$, so that the Markov chains generated by A and its reversed process are identical. In particular, equality holds when A is a tridiagonal Markov chain matrix.*

Proof. For (2.14a), take $B = A'$ in (2.5) and note that $\rho(A' \circ A') = \rho(A \circ A)$. To achieve (2.14b) we specify $B = S'AS = S^{-1}AS$ and check that $\rho(S^{-1}AS \circ S^{-1}AS) = \rho(A \circ A)$, since $S^{-1}AS = \|a_{\sigma(i),\sigma(j)}\|$ for an appropriate permutation σ of $\{1,2,\dots,m\}$ to itself. ■

The conjunction of (2.14a) and (2.14b) produces

$$\rho(A \circ S'A'S) \leqslant \rho(A \circ A). \tag{2.15}$$

The following convexity and monotonicity properties follow with the aid of (2.5).

THEOREM 2.2. *Consider a nonnegative irreducible matrix P and its Schur powers $P^{[r]}$. Let $\rho_r = \rho(P^{[r]})$. Then for all positive r,*

$$\rho_r^2 \leqslant \rho_{r-1}\rho_{r+1}, \tag{2.16}$$

so that $\log \rho_r$ *is convex. (When P is a strictly positive matrix, then* (2.16) *holds for all real r.) Strict inequality applies unless $P = c\tilde{U}$, where $\tilde{U}$ is any irreducible matrix of all unit or zero entries and c is a positive constant. Moreover,*

$$\frac{\log \rho_r}{r} \text{ strictly decreases for } r > 0. \tag{2.17}$$

When P is an irreducible stochastic matrix distinct from $c\tilde{U}$, then

$$\frac{\log \rho_r}{r-1} \text{ strictly increases over integer } r \geqslant 2. \tag{2.18}$$

Proof of Theorem 2.2. The proof of (2.16) follows from Theorem 2.1 by setting $A = P^{[(r+1)/2]}$, $B = P^{[(r-1)/2]}$. Alternatively, we could also deduce (2.16) (except for the matter of strictness) by appealing to a general result of Kingman [7] to the effect that when $P(\theta) = \|p_{ij}(\theta)\|$ is a nonnegative matrix depending on a real (or multivariate) parameter θ, where each term is log convex in θ over a domain Δ, then $\log \rho(P(\theta))$ is convex in θ on the same domain. Obviously in the matrix $P^{[r]} = \|p_{ij}^r\|$, $p_{ij}^r = \exp[r \log p_{ij}]$ is log convex as a function of r, and (2.16) ensues.

The function $f(r) = \log \rho_r$ is also subadditive, i.e.,

$$f(r+s) \leqslant f(r) + f(s), \tag{2.19}$$

on account of (2.4). The inequality in (2.19) is strict when P is irreducible nonnegative. Combination of (2.16) and (2.19) entails (2.17), as we now show.

Consider the positive values r and s obeying $r < s < 2r$ such that

$$s = \frac{2r-s}{r} r + \frac{s-r}{r} 2r.$$

Convexity implies

$$f(s) \leqslant \frac{2r-s}{r} f(r) + \frac{s-r}{r} f(2r).$$

By subadditivity $f(2r) < 2f(r)$ and therefore

$$f(s) < \frac{2r-s}{r} f(r) + \frac{s-r}{r} 2f(r) = \frac{s}{r} f(r)$$

or $f(s)/s < f(r)/r$, as asserted in (2.17).

We now turn to the proof of (2.18). For P a stochastic matrix the inequality (2.16) becomes

$$\frac{\rho_2}{\rho_1} = \rho_2 \leqslant \frac{\rho_3}{\rho_2} \leqslant \cdots \leqslant \frac{\rho_{r+1}}{\rho_r} \leqslant \cdots. \tag{2.20}$$

Therefore

$$\prod_{i=1}^{r-1} \frac{\rho_{i+1}}{\rho_i} \leqslant \left(\frac{\rho_{r+1}}{\rho_r} \right)^{r-1},$$

which simplifies to

$$\rho_r \leqslant \left(\frac{\rho_{r+1}}{\rho_r} \right)^{r-1}, \qquad \text{or} \quad (\rho_r)^r \leqslant (\rho_{r+1})^{r-1},$$

showing that

$$(\rho_r)^{1/(r-1)} \text{ is increasing for integer } r = 2,3,\ldots,$$

as stated in (2.18). This monotonicity is strict provided $P \neq cU$, where $\tilde{U}$ is a matrix of all unit zero or entries. The proof of Theorem 2.2 is complete. ∎

In the special case where

$$P = \|p_j u_i\|, \qquad p_j > 0, \quad \sum_{j=1}^{m} p_j = 1, \quad u_i \equiv 1,$$

the inequalities (2.17) and (2.18) translate into the inequalities

$$\left(\sum_{i=1}^{m} p_i^r\right)^{1/r} \downarrow \quad \text{in } r \qquad \text{while} \quad \left(\sum_{i=1}^{m} p_i^r\right)^{1/(r-1)} \uparrow, \tag{2.21}$$

the latter monotonicity applying for integer $r \geqslant 2$.

REMARK 3. The next argument establishes the lower bound

$$\rho_r \geqslant \frac{1}{m^{r-1}} \tag{2.22}$$

for any stochastic matrix P. To this end, we apply Hölder's inequality to obtain

$$1 = \sum_{j=1}^{m} p_{ij} \leqslant \left(\sum_{j=1}^{m} p_{ij}^r\right)^{1/r} \left(\sum_{j=1}^{m} 1\right)^{(r-1)/r}$$

and therefore

$$\frac{1}{m^{r-1}} \leqslant \sum_{j=1}^{m} p_{ij}^r \qquad \text{for all } i. \tag{2.23}$$

The relation (2.23) can be expressed as

$$P^{[r]}\mathbf{u} \geqslant \frac{1}{m^{r-1}}\mathbf{u}, \qquad \mathbf{u} = (1,\ldots,1). \tag{2.24}$$

Cognizance of the spectral radius characterization (2.1) implies that

$$\rho_r = \rho(P^{[r]}) \geqslant \frac{1}{m^{r-1}}. \tag{2.25}$$

Equality occurs if and only if $p_{ij} \equiv 1/m$.

Motivated by the probability model for generating successive letters following a Markov chain as compared to independent letter determinations, and in view of (2.25), it is tempting to surmise the inequality $\rho_r(P) \geqslant \rho_r(P_\infty)$ with $P_\infty = \lim_{k\to\infty} P^k$. This inequality is generally not correct. Indeed, take

$$P = \begin{pmatrix} .78 & .22 \\ .88 & .12 \end{pmatrix}, \qquad P_\infty = \begin{pmatrix} .8 & .2 \\ .8 & .2 \end{pmatrix};$$

then

$$\rho_2(P) = 0.666, \qquad \rho_2(P_\infty) = 0.68,$$

while for

$$Q = \begin{pmatrix} .82 & .18 \\ .72 & .28 \end{pmatrix}, \qquad Q_\infty = \begin{pmatrix} .8 & .2 \\ .8 & .2 \end{pmatrix}$$

we get

$$\rho_2(Q) = 0.699, \qquad \rho_2(Q_\infty) = 0.68.$$

REMARK 4. The natural functional extension of Schur powers with respect to a positive function $\varphi(\cdot)$ is the matrix

$$P^{[\varphi]} = \left\| \varphi(p_{ij}) \right\|. \tag{2.26}$$

The generalized Schur products involving two functions φ and ψ are denoted by

$$P^{[\varphi]} \circ Q^{[\psi]} = \left\| \varphi(p_{ij}) \psi(q_{ij}) \right\|.$$

It would be of interest to investigate the following problem. Let P and Q be nonnegative matrices, and let φ, ψ, and θ defined on the nonnegative axis be strictly increasing functions. Under what conditions [cf. (2.13)] does

$$\theta(\rho(P \circ Q)) \leqslant \varphi^{-1}(\rho(P^{[\varphi]}))\psi^{-1}(\rho(Q^{[\psi]})) \tag{2.27}$$

hold?

There are versions of Theorems 2.1 and 2.2 in terms of metrics on Orlicz spaces and extensions to Schur products of completely continuous integral kernel operators. These will be presented elsewhere.

3. RESULTS ON SCHUR POWERS IN TERMS OF ENTROPY AND MOST PROBABLE CONFIGURATIONS

A. *Schur Powers and Entropy*

The following theorem establishes a connection between the entropy of a finite stationary Markov chain and $d\rho_r/dr|_{r=1}$ (see Theorem 2.2).

THEOREM 3.1. *Let P be an irreducible Markov chain matrix with stationary frequency vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$, i.e., $\boldsymbol{\pi}P = \boldsymbol{\pi}$. Let $\rho_r = \rho(P^{[r]})$. Then*

$$\left.\frac{d\rho_r}{dr}\right|_{r=1} = H(\{X_n\}), \tag{3.1}$$

where H is the entropy of the Markov chain $\{X_n\}$ associated with P (see e.g. [6, Chapter 9] for the definition of the entropy of a stationary Markov chain).

It is known (e.g., [6, p. 497]) that

$$H(\{X_n\}) = \sum_{j,i=1}^{m} \pi_i p_{ij} \log p_{ij}.$$

Proof. Let $\mathbf{z}(r) = (z_1(r), z_2(r), \dots, z_m(r))$ be the unique principal eigenvector for $P^{[r]}$ corresponding to the principal eigenvalue ρ_r normalized so that

$$\langle \mathbf{z}(r), \mathbf{u} \rangle = \sum_{i=1}^{m} z_i(r) = 1.$$

($\langle \mathbf{z}, \mathbf{w} \rangle = \Sigma z_i w_i$ is the standard Euclidean inner product.) Since P is irreducible nonnegative, it is elementary that $\mathbf{z}(r)$ is an analytic function of r.

Differentiating

$$P^{[r]}\mathbf{z}(r) = \rho_r \mathbf{z}(r)$$

in r gives (we write $\dot{z}_j(r) = dz_j(r)/dr$)

$$\sum_j (p_{ij})^r \dot{z}_j(r) + \sum_j (\log p_{ij}) p_{ij}^r z_j(r) = \frac{d\rho_r}{dr} z_i(r) + \rho_r \dot{z}_i(r).$$

Setting $r = 1$ and forming the inner product with $\boldsymbol{\pi}$, since $\boldsymbol{\pi}P = \boldsymbol{\pi}$, $\rho_1 = 1$, noting that $z_j(1) \equiv 1/m$, $j = 1, \ldots, m$, we obtain

$$\sum_i \pi_i \sum_{j=1}^{m} p_{ij} \dot{z}_j(1) + \frac{1}{m} \sum_i \pi_i \sum_j (\log p_{ij}) p_{ij} = \left.\frac{d\rho_r}{dr}\right|_{r=1} \left(\sum_i \pi_i\right) \cdot \frac{1}{m} + \sum_i \pi_i \dot{z}_i(1),$$

which reduces to

$$\sum_j \pi_j \dot{z}_j(1) + \frac{1}{m} H(\{X_n\}) = \frac{1}{m} \cdot \left.\frac{d\rho_r}{dr}\right|_{r=1} + \sum_i \pi_i \dot{z}_i(1)$$

and therefore

$$H(\{X_n\}) = \sum_{i,j} \pi_i p_{ij} \log p_{ij} = \left.\frac{d\rho_r}{dr}\right|_{r=1}$$

as required to be shown. ■

B. Minimum Spectral Radius of Schur Powers

It is of interest to characterize $\lim_{r\to\infty} \rho_r^{1/r} = \inf_{r>0} \rho_r^{1/r}$; see Theorem 2.2. Consider P as a stochastic (Markov) matrix.

We claim that the lower bound of $\inf_{r>0}(\rho_r)^{1/r}$ is the limiting geometric mean of the most likely probability over all sample paths of the Markov chain process with transition probability matrix P. Consider a sequence of states (letters) generated as a realization following the Markov chain matrix $\mathbf{P} = \|p_{ij}\|_1^m$. Let $p(w)$ be the probability of a $k+1$-length word w, indicating the outcome at the first $k+1$ time points. For example, if $w = \{i_1, i_2, \ldots, i_{k+1}\}$ the probability $p(w)$ is $\pi_{i_1} \prod_{\nu=1}^{k} p_{i_\nu i_{\nu+1}}$ where π_i is the probability that the initial state is i.

We can identify

$$\sum_{w \in \mathscr{W}_{k+1}} [p(w)]^r = \left\langle \boldsymbol{\pi}^{[r]}, (P^{[r]})^k \mathbf{u} \right\rangle$$

(w_{k+1} denotes the collection of all $k+1$-length words) where $(P^{[r]})^k$ is the kth matrix product of the rth Schur power $P^{[r]}$, $\pi^{[r]}=(\pi_1^r,\pi_2^r,\dots,\pi_m^r)$, and $\mathbf{u}=(1,\dots,1)$. Consider the expansion $(P^{[r]})^k\xi=(\rho_r)^k\langle\psi,\xi\rangle\varphi+O(\gamma_{[r]}^k)$, where φ (ψ) is the right (left) principal eigenvector of $P^{[r]}$, normalized so that $\langle\varphi,\psi\rangle=1$. From the theory of positive matrices we know $|\gamma_{[r]}|<\rho_r$. Obviously

$$\lim_{k\to\infty}\Big(\sum_{w\in\mathscr{W}_{k+1}}[p(w)]^r\Big)^{1/k}=\lim_{k\to\infty}\Big\langle\pi^{[r]},(P^{[r]})^k\mathbf{u}\Big\rangle^{1/k}=\rho_r.$$

Observe that

$$\begin{aligned}\sum_{w\in\mathscr{W}_{k+1}}[p(w)]^r&\leqslant\max_{w\in\mathscr{W}_{k+1}}\big([p(w)]^{r-1}\big)\sum_{w\in\mathscr{W}_{k+1}}p(w)\\&=\max_{w\in\mathscr{W}_{k+1}}[p(w)]^{r-1}\end{aligned}$$

and so for any fixed $r>0$

$$\rho_r^{1/r}=\left[\lim_{k\to\infty}\Big(\sum_{w\in\mathscr{W}_{k+1}}[p(w)]^r\Big)^{1/k}\right]^{1/r}\leqslant\lim_{k\to\infty}\Big(\max_{w\in\mathscr{W}_{k+1}}[p(w)]\Big)^{(r-1)/r-k},$$

and therefore

$$\inf_{r>0}\rho_r^{1/r}=\lim_{r\to\infty}\rho_r^{1/r}\leqslant\varliminf_{k\to\infty}\Big(\max_{w\in\mathscr{W}_{k+1}}[p(w)]\Big)^{1/k}. \qquad (3.2)$$

To reverse the inequality, with any $\varepsilon>0$ we take r_0 fixed so that

$$\inf_{r>0}\rho_r^{1/r}+\varepsilon>\rho_{r_0}^{1/r_0}=\lim_{k\to\infty}\Big(\sum_{w\in\mathscr{W}_{k+1}}[p(w)]^{r_0}\Big)^{1/kr_0}$$

and, keeping a single term of the sum,

$$\geqslant\varlimsup_{k\to\infty}\Big(\max_{w\in\mathscr{W}_{k+1}}[p(w)]\Big)^{1/k}. \qquad (3.3)$$

The combination (3.2) and (3.3) proves the identity

$$\lim_{r\to\infty}\rho_r^{1/r}=\lim_{k\to\infty}\Big(\max_{w\in\mathscr{W}_{k+1}}[p(w)]\Big)^{1/k}, \qquad (3.4)$$

displaying the right hand side as the most likely realization (in the geometric sense) among the possible sample paths of the process.

We illustrate these ideas concretely for a two state Markov chain. Consider the example

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}.$$

Then it is easily checked that

$$\lim_{r \to \infty} \rho_r^{1/r} = \max\left(\alpha, \beta, \sqrt{(1-\alpha)(1-\beta)}\right)$$

and

when $\max = \alpha$, the most likely realization is $\{1,1,1,1,\ldots\}$ (repeated occurrence of state 1),

when $\max = \beta$, it is $\{2,2,2,2,\ldots\}$,

when $\max = \sqrt{(1-\alpha)(1-\beta)}$, it is $\{2,1,2,1,2,1,\ldots\}$

—all in the sense of (3.4).

Another useful simulation of ρ_r is encompassed by the formula

$$\rho_r = \lim_{k \to \infty} \Pr\{X^{(r)} \geqslant k+1 | X^{(r)} \geqslant k\}$$

$$= \lim_{k \to \infty} \frac{\Pr\{X^{(r)} \geqslant k+1\}}{\Pr\{X^{(r)} \geqslant k\}} = \lim_{k \to \infty} \frac{\sum_{w \in \mathscr{W}_{k+1}} [p(w)]^r}{\sum_{w \in \mathscr{W}_k} [p(w)]^r}, \tag{3.5}$$

where $X^{(r)}$ counts the number of matchings of r independent strings of state values each generated by the Markov chain matrix P. We omit the formal proof.

The analysis leading to (3.4) entails the following uniform exponential bounds. For any fixed r there exists a $\gamma < \rho_r^{1/r}$ and a constant C depending only on r such that

$$p_{\mu j_1} p_{j_1 j_2} \cdots p_{j_k \nu} \leqslant C\gamma^k$$

for all specifications of nonnegative integers $\mu, j_1, \ldots, j_k, \nu$.

4. MAJORIZATION COMPARISONS FOR THE SPECTRAL RADIUS OF SCHUR POWERS

The motivation stems from the following example. Let $P = \|u_i p_j\|_1^m$ be the rank one stochastic matrix ($u_i \equiv 1$, $p_j > 0$, $\Sigma p_j = 1$) and $Q = \|u_i q_j\|_1^m$, where

$$\mathbf{q} = T\mathbf{p} \tag{4.1}$$

and $T = \|t_{ij}\|$ is doubly stochastic, that is, $T \geqslant 0$, $T\mathbf{u} = \mathbf{u}$, and $\mathbf{u}T = \mathbf{u}$, so that the frequency vector $\mathbf{p}$ majorizes $\mathbf{q}$. (For extensive references on the notion of majorization of vectors see [8].) Since $\rho(P^{[r]}) = \Sigma_{j=1}^m p_j^r$, we have the inequality

$$\rho(Q^{[r]}) \leqslant \rho(P^{[r]}) \qquad \text{for every integer } r \tag{4.2a}$$

as a simple consequence of the majorization inequality $\Sigma_j q_j^r \leqslant \Sigma_j p_j^r$. More generally, since for any convex function $\varphi(\cdot)$ we have $\Sigma_{j=1}^m \varphi(q_j) \leqslant \Sigma_{j=1}^m \varphi(p_j)$ for $\mathbf{q}$ and $\mathbf{p}$ related as in (4.1), we deduce

$$\rho(Q^{[\varphi]}) \leqslant \rho(P^{[\varphi]}) \tag{4.2b}$$

[see (2.26) for the definition of $P^{[\varphi]}$].

Traditional multivariate majorization orderings relative to matrices generalizing (4.1) are as follows. A matrix P is said to *majorize the matrix* Q if there exists a doubly stochastic matrix T fulfilling the equation

$$Q = PT \qquad (\text{symbolized by} \quad P \succ Q) \tag{4.3}$$

(cf. [5]). When P is stochastic, Q is also stochastic.

In comparing stochastic matrices a symmetrical notion of majorization requires that there exist a doubly stochastic matrix T connecting P and Q by the relation

$$Q = T'PT \qquad \left(\text{symbolized by} \quad P \overset{s}{\succ} Q\right). \tag{4.4}$$

We inquire, in line with (4.2), for Q and P satisfying (4.3) or (4.4), whether

$$\rho(Q^{[r]}) \leqslant \rho(P^{[r]}). \tag{4.5}$$

When T is a permutation matrix the fulfillment of (4.5) mandates equality. However, for

$$T=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then for the model (4.3), taking

$$P=\begin{pmatrix} \alpha & 1-\alpha \\ 1 & 0 \end{pmatrix} \qquad \text{with} \quad 0<\alpha<1$$

and accordingly

$$Q=\begin{pmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{pmatrix},$$

we obtain, for $r>1$, $\rho(Q^{[r]})=1$ while $\rho(P^{[r]})<1$.

Subject to the relation (4.4) for any permutation matrix T, we have $\rho(Q^{[r]})=\rho(P^{[r]})$.

THEOREM 4.1. *If P is a symmetric stochastic matrix and $Q \overset{s}{\prec} P$, then for any positive convex function $\varphi(\cdot)$*

$$\rho(Q^{[\varphi]}) \leqslant \rho(P^{[\varphi]}).$$

Proof. By convexity we have

$$\varphi(q_{ij})=\varphi\Big(\sum_{k,l} t_{ki}p_{kl}t_{lj}\Big) \leqslant \sum_k t_{ki}\varphi\Big(\sum_l p_{kl}t_{lj}\Big) \leqslant \sum_{k,l} t_{ki}\varphi(p_{kl})t_{lj},$$

or componentwise $Q^{[\varphi]} \leqslant T'P^{[\varphi]}T$. By virtue of the characterization (2.1), we have

$$\rho(B) \leqslant \rho(T'AT), \tag{4.6}$$

where $A=P^{[\varphi]}$ and $B=Q^{[\varphi]}$.

Since A is a symmetric matrix (without loss of generality we can assume T nonsingular) we have

$$\begin{aligned}\rho(T'AT) &= \sup_{\mathbf{x}\neq 0}\frac{\langle T'AT\mathbf{x},\mathbf{x}\rangle}{\langle \mathbf{x},\mathbf{x}\rangle} \leqslant \sup_{\mathbf{x}\neq 0}\frac{\langle AT\mathbf{x},T\mathbf{x}\rangle}{\langle T\mathbf{x},T\mathbf{x}\rangle}\sup_{\mathbf{x}\neq 0}\frac{\langle T\mathbf{x},T\mathbf{x}\rangle}{\langle \mathbf{x},\mathbf{x}\rangle} \\ &\leqslant \rho(A)\rho(T'T)=\rho(A), \end{aligned} \tag{4.7}$$

since T is doubly stochastic. The conjunction of (4.6) and (4.7) proves the desired result. ■

The method of Theorem 4.1 also proves

THEOREM 4.2. *Let $Q=PT$ where P is stochastic and symmetric. Let T be doubly stochastic and symmetric. If either T or P is positive definite, then for any $\varphi(\xi)=\Sigma_{k=0}^{\infty}a_k\xi^k$, for $0\leqslant\xi\leqslant 1$, $a_k\geqslant 0$, $\Sigma a_k>0$,*

$$\rho(Q^{[\varphi]})\leqslant\rho(P^{[\varphi]}).$$

Under the conditions of the theorem it follows, if $A=P^{[\varphi]}$ is positive definite whenever P is, that

$$\begin{aligned}\rho(AT)&=\sup_{\mathbf{x}\neq 0}\frac{\langle A^{1/2}TA^{1/2}\mathbf{x},\mathbf{x}\rangle}{\langle\mathbf{x},\mathbf{x}\rangle}\\&\leqslant\sup_{x\neq 0}\frac{\langle TA^{1/2}\mathbf{x},\,A^{1/2}\mathbf{x}\rangle}{\langle A^{1/2}\mathbf{x},\,A^{1/2}\mathbf{x}\rangle}\sup_{\mathbf{x}\neq 0}\frac{\langle A\mathbf{x},\mathbf{x}\rangle}{\langle\mathbf{x},\mathbf{x}\rangle}\\&=\rho(A).\end{aligned}$$

When T is positive definite, then it suffices to have $\varphi(\xi)$ convex and positive over $0\leqslant\xi\leqslant 1$.

REMARK. The conclusion of Theorem 4.1 prevails whenever $\rho(T'AT)\leqslant\rho(A)$. However, this inequality can fail for A merely nonnegative even for *second order* matrices. Indeed, consider

$$A=\begin{pmatrix}a&b\\c&d\end{pmatrix}\quad\text{and}\quad T=\begin{pmatrix}\alpha&1-\alpha\\1-\alpha&\alpha\end{pmatrix}.$$

The characteristic polynomial of $T'AT$ is $\lambda^2-\lambda[(a+d)+2\alpha(1-\alpha)(b+c-a-d)]+(ad-bc)(2\alpha-1)^2=0$. For $ad-bc=0$ we can have $\rho(T'AT)$ either increasing or decreasing as α deviates from $\frac{1}{2}$, depending on the sign of $a+d-b-c$.

Another more direct comparison among Schur products and averages thereof considers a collection $A_1, A_2,\ldots, A_r$ of nonnegative matrices of order

m and the construction

$$A'_\nu = \sum_{\mu=1}^{r} t_{\nu\mu} A_\mu, \qquad \nu = 1,2,\dots,r,$$

where $T = \|t_{ij}\|$ is doubly stochastic. We claim that

$$\rho\left(A'_1 \circ A'_2 \circ \cdots \circ A'_r\right) \geqslant \rho(A_1 \circ A_2 \circ \cdots \circ A_r), \tag{4.8}$$

where for each fixed i, j, we write $A_\nu = \|a_{ij;\,\nu}\|$.

In fact, since $\{a'_{ij;\,\nu}\}_{\nu=1}^r$ is majorized by $\{a_{ij;\,\nu}\}_{\nu=1}^r$, we have

$$\prod_{\nu=1}^{r} a'_{ij;\,\nu} \geqslant \prod_{\nu=1}^{r} a_{ij;\,\nu},$$

so that componentwise

$$A'_1 \circ A'_2 \circ \cdots \circ A'_r \geqslant A_1 \circ A_2 \circ \cdots \circ A_r,$$

and therefore (4.8) obtains.

The analog of (4.8) would apply for any Schur concave matrix function comparing $\rho[\varphi(A'_1, A'_2,\dots, A'_r)]$ with $\rho[\varphi(A_1, A_2,\dots, A_r)]$.

REFERENCES

1 J. E. Cohen, Random evolutions and the spectral radius of a nonnegative matrix, *Proc. Cambridge Philos. Soc.* 86:345–350 (1979).
2 C. H. Fitzgerald and R. A. Horn, On the structure of Hermitian symmetric inequalities, *J. London Math. Soc.* 15:419–430.
3 N. A. Kahn and A. Marcus, A note on the Hadamard product, *Canad. Math. Bull.* 2:81–83 (1959).
4 S. Karlin and F. Ost, Maximal segmental match length among random sequences from a finite alphabet, in *Volume in Honor of Neyman and Kiefer, Berkeley Symposium* (L. LeCam, Ed.), to appear.
5 S. Karlin and Y. Rinott, Comparison of measures, multivariate majorization, and applications in statistics, in *Studies in Econometrics, Time Series and Multivariate Statistics* (S. Karlin, T. Amemiya, and L. A. Goodman, Eds.), Academic, San Diego, 1983, pp. 465–489.

6 S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, Academic, New York, 1975.
7 J. F. C. Kingman, A convexity property of positive matrices, *Quart. J. Math. Oxford Ser. (2)* 12:283–284 (1961).
8 A. W. Marshall and I. Olkin, *Theory of Majorization Inequalities and Its Applications*, Academic, New York, 1979.
9 Q. F. Stout, Schur multiplication on $\mathscr{B}(l_p, l_q)$, *J. Operator Theory* 5:231–243 (1981).
10 G. P. H. Styan, Hadamard products and multivariate statistical analyses, *Linear Algebra Appl.* 6:217–240 (1973).

Received 27 October 1983; revised 5 April 1984